DOCUMENT RESUME

ED 053 134 24 TE 002 525

AUTHOR	Lee, Richard R.
TITLE	Effects of Age on Student Perception of Social Dialects. Final Report.
INSTITUTION	Florida State Univ., Tallahassee. Dept. of Speech.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
BUREAU NO	BR-9-D-021
PUB DATE	Jun 71
GRANT	OEG-4-9-190021-0032-057
NOTE	12p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*Age Differences, *Connected Discourse, Interviews, *Language Usage, *Nonstandard Dialects, *Socioeconomic Influences, Testing

ABSTRACT

This study was built on the hypothesis that linguistic cues distinguish the speech of the poor and the ethnically different. Eight samples of connected interview discourse were taken from a two-by-two-by-two matrix of sex, race and social status. The samples were played to groups of elementary, secondary and college students who rated each voice for the highest job each speaker could hold. Listeners agreed strongly about the relative quality of speech demanded by each occupation, but they did not agree with one another in their ratings of the voices or with themselves when they were retested. This study does not support the viability of dialect as a reliable cue in social perception. (Author)

BR 9-D-021
PA 24
TE

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED053134

FINAL REPORT

PROJECT NO. 9-D-021
GRANT NO. OEG-4-9-190021-0032-057

Richard R. Lee
DEPARTMENT OF SPEECH
FLORIDA STATE UNIVERSITY
Tallahassee, Florida 32306

EFFECTS OF AGE ON STUDENT PERCEPTION OF SOCIAL DIALECTS

June 1971

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Office of Education

National Center for Educational Research and Development
Regional Research Program

TE 002 525

ABSTRACT

This study was built on the hypothesis that linguistic cues distinguish the speech of the poor and the ethnically different. Eight samples of connected interview discourse were taken from a two-by-two-by-two matrix of sex, race and social status. The samples were played to groups of elementary, secondary and college students who rated each voice for the highest job each speaker could hold. Listeners agreed strongly about the relative quality of speech demanded by each occupation but they did not agree with one another in their ratings of the voices or with themselves when they were retested. This study does not support the viability of dialect as a reliable cue in social perception.

FINAL REPORT

PROJECT NO. 9-D-021
GRANT NO. OEG-4-9-190021-0032-057

EFFECTS OF AGE ON STUDENT PERCEPTION OF SOCIAL DIALECTS

Richard R. Lee

FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA

18 June 1971

The research reported herin was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education

INTRODUCTION

One of the working assumptions among professional organizations of English teachers is that the speech of lower class speakers is a barrier to social mobility. In his presidential address to the Teachers of English as a Second Language in 1970, David P. Harris noted the shifting attention of the organization and a growing concern "for those many thousands of American children and adults whose academic success and social mobility are severely restricted by the kind of English they use."[1] This concern has been translated into a special kind of instruction for the children of the poor instruction called dialect modification or standard English as a second language. The postulate which justifies such instruction has never been adequately tested, that is, that the speech of members of the lower classes constitutes a barrier. Implicit in Harris' observation are several assumptions about the relationship between speech and social perception which must be made explicit before the barrier postulate can be tested. The first assumption is that differences exist in the speech of different social classes in America which are best described as differences in dialect, specifically differences in phonology and syntax. One compilation of these phonological and syntactic differentia has been made by Raven McDavid [2] It should be noted in passing that there are many other ways to describe speech differences: sentence length, word choice, type token ratios, appropriateness of responses, and so on. A second assumption is that untrained listeners detect and isolate these phonological and syntactic differentia within the message, a detection which negatively influences the listener's evaluation of the social worth of the speaker. Stated another way, this assumption says that how a speaker says something is perceived independently of what he says and the how has priority in the process of forming social judgments. A third assumption is that these differentia provide a reliable cue to a speaker's social class for the casual listener. In other words, 30 listeners acting independently should all come to about the same conclusion about a speaker's rank in the social hierarchy. Another assumption is that speech differentia are common to members of a social class or ethnic minority. Inherent in the barrier postulate is yet another assumption, that for the average listener, the identification of a speaker's social class is a primary percept, a cognition formed early in an encounter. These assumptions constitute the barrier postulate and provide a rationale for instruction in dialect modification.

The research by which such teaching is justified is inconclusive. Putnam and O'Hearn reported a strong association between a speaker's real social class and listeners judgment about their social class after listening to their speech.[3] Their dependent measure, however, made the purposes of their experiment transparent and thus may have been reactive. Harms, in an effort to control for content, had speakers reply to questions and directions printed on cue cards.[4] Each listener then heard 20 or so repetitions of the same messages in succession. His study raises the question of ecological validity for the listeners. In another content-controlled experiment Tucker and Lambert[5] had many subjects read the same passage, which again raises the problem of the validity of the listening situation. In addition, Tucker and Lambert drew their conclusions by ranking mean scores, apparently without testing for dispersion. Labov's[6] listeners rated content-controlled samples spoken by others after they read an identical passage aloud. Labov's dependent measure, a rating of occupations on the base of occupational prestige, was not tested for consensual validity.

To recapitulate, research in the field of dialect perception is open to question on grounds of ecological validity because raters listen to the same message repeatedly. The construction of each of these experiments leads one to suspect experimental reactivity. None of the dependent measures have been tested for consensual validity. Finally, these experiments have been interpreted without subjecting the data to tests of interrater and intrarater reliability

METHOD

Speech Sampling Procedure

In this study speech samples were collected from interviews conducted with eight subjects between 15 and 17 years old, four boys and four girls, four blacks and four whites, and four from families making more than $5000 a year and four speakers from families earning less than $3000 a year. To control reasonal variations in the sample, all subjects were born and reared within 25 miles of Tallahassee, Florida. Notice that the independent variables are race, sex, and socio-economic class, not dialect.

Each speaker talked with an interviewer who asked what the individual wanted to do for a living, how he spent his free time,

what his friends were like, what he looked for in a job, and what his interests were. No two speakers were asked the same set of questions. After the interview, each speaker read a short narrative passage into the tape recorder. Two versions of each interview were prepared. In the unedited version, his two or three most fluent replies along with the antecedent question were lifted from the interview. These excerpts were spliced together to form the unedited sample. Each individual's sample was 45 to 60 seconds long. From these unedited samples, another, shorter, version was derived by eliminating all non-linguistic extrania such as filled pauses, bad sentence starts and silences of more than one second. These edited tapes contained exactly the same words as the unedited versions except that each speaker sounded more fluent.

PROCEDURES, PILOT EXPERIMENT (ELEMENTARY SCHOOL)

A preliminary version of the dependent measure was developed by taking a class of 10 7-year-olds from a second grade school. The first question is whether this racially mixed group had any notions about the speech requirements of different occupations. They did. What's more, they could distinguish between occupational status and speech qualification. In the first preliminary version, they were given sketches on 3x5 cards of a teacher, a television announcer, a doctor, an office worker, an artist, a housewife, a truck driver and a store clerk and asked to rank the cards on the basis of which occupations demanded the best speech. These second graders regularly ranked the artist at the bottom of the scale, that is, where "it doesn't make any difference how you talk." In a second preliminary version, the artist was removed and the occupation of "truck driver" and "housewife" was collapsed into a single category because these were regularly ranked sixth and seventh, and because the stimulus tapes had both male and female speakers.

RESULTS, PILOT EXPERIMENT (ELEMENTARY SCHOOL)

This class of second graders finally ranked these occupations on a continuum from those requiring the best speech to those occupations having no speech qualifications at all.

1. announcer
2. teacher
3. doctor
4. office worker
5. store clerk
6. truck driver or housewife

The Kendal co-efficient of concordance for these rankings within the group was .49.

The next step was to have the second graders rate the eight speech samples by this dependent measure. The edited samples were played to each child in individual session. After each voice, the child was asked to indicate the highest job that the speaker could hold on the basis of the way he talked. These ratings were then subjected to Ebel's formula for individual inter-rater reliability. The co-efficient for the second graders was minus .03. (A minus co-efficient is possible if the error term is monumental). The second graders were retested on the occupational ranking and speech rating tasks two weeks later. Individual inter-rater reliability of the speech rating was .00. Individual protocols for the test and the re-test were compared. Virtually every child's rating of every voice was different by at least one rank and often by three or four. The absence of inter-rater reliability and a visual scan of intra-rater reliability suggested that the speech rating activity was guessing behavior.

PROCEDURES AND RESULTS, SECOND PILOT EXPERIMENT (SECONDARY SCHOOL)

It was felt that two factors may have contributed to the instability of these ratings: the mixed racial makeup of the second grade class with the resulting heterogeneity of social backgrounds and their tender age. Nineteen ninth-graders, ages 13 to 15, were selected from an all black school. All students had records of frequent brushes with school authority. Five of the nineteen listeners were on county assistance, and all but one of the remaining students had fathers in unskilled or semi-skilled occupations. In individual sessions each subject ranked the six occupations on the basis of the speech qualifications demanded

of each. The rankings for the group rendered a Kendall co-efficient of concordance of .77. Each of the nineteen listeners rated the eight edited speech samples. The individual inter-rater reliability of these ratings was .08. Homogeneity of race and social background, even among these adolescents, produced no concensus when assigning occupational potential to eight real real voices.

An older and academically more homogenous group was selected from nineteen white seniors ages 17 through 19, in a small selective high school with a long waiting list. All of these students planned to enter college in a few months. Individuals again ranked the six occupations on a continuum from those that demanded the best speech, those which had no speech qualification at all. The Kendall co-efficient of concordance was .89. Individual ratings of the speech qualification of the eight edited voice samples was tested for individual inter-rater reliability, the test produced a co-efficient of .18. That is not what one would call concensus.

At least one trend is apparent in the data reported so far. As a group, second graders seem to have a fairly stable idea of which jobs require the best speech. Group concurrence seems to increase to near unanimity as a function of age level. Utilizing this knowledge in passing judgments about voices, however, is another matter. If the dialect of the poor and the black act as a barrier, it is a wholly unreliable barrier among these listeners.

PROCEDURES, THIRD EXPERIMENT (COLLEGE)

Two more steps were taken to try to find some coherence in the data. First, the scale was reduced from six categories to four; announcer, teacher, office worker and the composite truck driver/housewife. The reasoning was that perhaps every listener did not have six perceptual or judgmental categories. The test was also moved up among college students to see if a more homogenous, better educated, older population might show a moderate amount of inter-rater reliability.

It should be noted that no experiment in dialect perception had ever used free conversion as a stimulus. It would be interesting to test the influence of edited free speech, its unedited equivalent,

and the content-controlled samples taken when the speakers recorded the narrative passage. In the following experiment all three sample types were used. In the test, listeners heard all eight edited samples, then the eight narratives, and then the eight unedited samples. The order of the speakers was rerandomized for all three conditions. After each presentation each listener rated the highest job the speaker could hold on the basis of his speech. The listeners returned a week later and repeated the same task with the order of conditions reversed. At the end of the tetest section the listeners were asked to write down what they thought the experiment was all about.

RESULTS, THIRD EXPERIMENT (COLLEGE)

Consentual validity about the ranks of the rating scale was nearly perfect; only two of the 50 college raters did not follow the order of announcer - teacher - store-clerk - truck-driver/housewife.

The ratings from the test and the retest for each condition, edited and unedited interview, and the content-controlled reading passage, were tested by Ebel's formula for individual inter-rater reliability. The co-efficients are presented in Table I.

TABLE I

CO-EFFICIENT OF INDIVIDUAL INTER-RATER RELIABILITY

	Test	Retest
Unedited	.52	.56
Edited speech	.30	.50
Reading of passage	.48	.58

To guage intra-rater reliability each raters score, for each of the eight speakers voices, under each of three conditions, in the test was compared with the score he gave the same voice, under the same condition a week later. There is no convenient statistic for this, so tied scores were counted. Of a possible 1200 (50 raters, 8 voices, 3 conditions), 655 were ties, 178 in the edited condition, 240 in the unedited, and 237 in the ratings given to the read passages.

Pure chance alone would account for 300 of the total 655 tied scores. By rater sex, men gave the same score on the retest 324 times, women 331. A rater-by-speaker data plot suggests no further patterning.

To check for experimental reactivity, 41 of the protocols were examined for free responses to the questions: What do you think this experiment is all about? Thirty-one listeners identified race as an experimental variable, and 17 identified social class.

DISCUSSION

This experiment speaks only ambiguously about the dialects of social classes and ethnic groups as a perceptual cue in social cognition. If it is a cue, it is not reliable, as evidenced by fluctuations within groups and within individuals over time. Individual inter-rater reliability rises from a minus value among the second-graders to a range between .30 and .50 among the college population and seems to increase somewhat with practice, but even the highest of these values falls far short of accounting for even half of the variance. It is interesting to notice that inter-rater reliability is roughly the same for unedited speech samples and the reading of passages. Whatever the impact of listening to repeated messages upon ecological validity, both elicitation techniques render about the same level of group agreement. Conversely when edited speech is the stimulus, not even the college group can agree among themselves or with themselves over time to any meaningful extent. What is suggested here is that non-linguistic cues, the pauses and bad starts edited out of the interviews, contribute substantially to what little stability was found in rating unedited speech and the reading of passages.

The 31 college listeners who identified race as a variable in the experiment and the 17 who identified social class, raise serious questions about the external validity of this experiment. Their sensing of the experimenter's purposes undoubtedly created a perceptual set which is uncharacteristic of behavior outside the laboratory. Despite this, neither the group nor the individuals within it could assign voice ratings with any degree of reliability worth mentioning.

No doubt the ambiguity of the dependent measure itself contributed to unreliability, yet any experimenter is caught between two poles: a need for an unobstrusive measure that doesn't tip his hand and the need for an unambiguous measure that does not add confounding factors of its own.

Nevertheless, in job interviews, in social transactions in the business community, and even in telephone conversations, a multitude of factors impinge on the listener, factors which are filtered and interpreted by individual characteristics of the listener. In the long view, it is a little presumptious to presume a nice, pat, linear relationship between one small aspect of the speech signal, and social perception by others. The perception of others is simply more complicated than that.

These findings suggest that it is pointless to pursue univariant studies between dialect and social perception further. An adequate design would have to take into account the characteristics of the listener, as well as the speaker. The controlled laboratory experiment at present seems ill suited to measure the dynamic, shifting flow of reaction and impression characteristic of diadic communication. The inert, passive evaluater has no counterpart in the real world. Communication usually takes place in a role-defined context, and always in context where speaker and listener have certain needs. The laboratory experiment negates both role and purpose.

If school instruction in dialect modification must be supported by emperical research, it will have to look further for justification.

References

1. David P. Harris, "1970 TESOL Presidential Report to the Membership," TESOL Newsletter, IV (September/December, 1970), p. 2.

2. Raven I. McDavid, Jr., "A Checklist of Significant Features for Discriminating Social Dialects," in Eldonna L. Evertts, Dimensions of Dialect, National Council of Teachers of English (Champaign, 1967), pp. 7-10.

3. George W. Putnam and Edna M. O'Hearn, "The Status Significance of an Isolated Urban Dialect," Language Supplement, Language Dissertation No. 53, XXXI, (October-December, 1955).

4. L. S. Harms, "Listener Judgments of Status Cues in Speech," Quarterly Journal of Speech, XLVII (1961), pp. 164-168.

5. Richard Tucker and Wallace E. Lambert, "White and Negro Listeners' Reactions to Various American-English Dialects," Social Forces, XLVII, (1969), pp. 463-468.

6. William Labov, The Social Stratification of English in New York City, Center for Applied Linguistics, (Washington, D. C., 1966).